Received: 17 October 2022 | Accepted: 24 January 2023 | Published online: 25 March 2023

DOI: 10.1002/jeq2.20458

Journal of Environmental Quality

TECHNICAL REPORT

Surface Water Quality

On the compounding of nitrate loads and discharge

Taereem Kim[1] | **Gabriele Villarini**[1] | **Hanbeen Kim**[1] | **Robert Jane**[2] | **Thomas Wahl**[2]

[1]IIHR—Hydroscience and Engineering, The University of Iowa, Iowa City, IA, USA

[2]Civil, Environmental, and Construction Engineering & National Center for Integrated Coastal Research, University of Central Florida, Orlando, FL, USA

Correspondence
Taereem Kim, IIHR—Hydroscience and Engineering, University of Iowa, 323 C. Maxwell Stanley Hydraulics Laboratory, Iowa City, IA 52242, USA.
Email: taereem-kim@uiowa.edu

Assigned to Associate Editor Yongshan Wan.

Abstract

Compound extremes can arise from combinations of multiple drivers, and even non-extreme univariate events can combine to cause large societal and economic impacts. In this study, we model multivariate compound events focusing on the potential interaction of nitrate loads and discharge. We use daily discharge and nitrate loads at seven US Geological Survey sites in the state of Iowa. We apply a two-sided conditional sampling method, which derives two joint probabilities conditioning on discharge and nitrate loads, respectively. Our results show that there is a dependence between discharge and nitrate loads, which can be described through bivariate modeling and the subsequent estimation of their joint annual exceedance probabilities (AEPs). The magnitude of the joint AEPs to extreme discharge and extreme nitrate loads exhibit different structures across the different sites, highlighting the different roles of these two quantities in controlling their compounding. In examining the ranges in design values for a given AEP, we found that the largest variability in highly likely values was generally associated with high agricultural intensity, high hog density, and fertilizer expenditures.

1 | INTRODUCTION

Compound extremes have been receiving growing attention and have been identified as a priority research area by the World Climate Research Program. Compound extremes can be largely grouped into four main categories: preconditioned, multivariate, spatially, and temporally compounding (Zscheischler et al., 2020). Among the multivariate compounding events, we can include, for example, compound flooding (e.g., riverine and coastal flooding), compound drought and heat (e.g., high temperature and low precipitation), humid heatwave (e.g., high temperature and atmospheric humidity), and compound precipitation and wind extremes (e.g., heavy precipitation and extreme winds) (Tilloy et al., 2021, 2020; Zscheischler et al., 2020). While there are growing efforts to examine the joint risk of compound events related to hydrologic variables such as floods (Sebastian et al., 2017; Wahl et al., 2015; Zellou & Rahali, 2019) and droughts (Manning et al., 2018; Ribeiro et al., 2019), one multivariate compound event that has received limited attention in the literature is the potential connection between water quantity and water quality. For instance, Oeurng et al. (2010) focused on a basin located in south-west France and examined 19 flood events; they found that nitrate transport was correlated with total precipitation, flood duration, peak discharge, and total water yield, while peak nitrate concentrations were not strongly controlled by peak discharge. Wang et al. (2017) developed a bivariate

model to relate discharge and nitrate/phosphorous to identify management strategies to mitigate water quality and quality issues of reservoir inflows. Liu et al. (2022) used a copula-based approach to develop an index that integrates water quality and quantity. Despite these efforts, the emphasis is generally not within the realm of a compound framework using multivariate extreme value analysis.

The natural hazards associated with water quantity and quality events have large societal and economic implications, especially in areas that have been experiencing a major intensification in agricultural practices. For example, since the 1940s the US Midwest has experienced a switch from perennial vegetation to seasonal row cropping, with large areas that have been intensively cultivated in corn and soybean (*Zea mays* and *Glycine max*). As a consequence of the changes in agricultural practices, nitrogen pollution has had major environmental and public health impacts. The Mississippi River Basin acts as the thread connecting the US Midwest to the Gulf of Mexico, where these large nitrogen loads cause algae blooms and the "Dead Zone" (Mueller & Helsel, 1996; Turner & Rabalais, 1994). From a human health perspective, the Environmental Protection Agency limits nitrate in drinking water to 10 mg/L because of issues related to "blue baby syndrome" with growing awareness of the chronic health effects for even lower concentrations over long exposure periods (Ward et al., 2018). Because of these issues, the Mississippi River/Gulf of Mexico Watershed Nutrient Task Force was established in 1997, which led to an action plan in 2008 for the establishment of nutrient reduction strategies for each of the 12 states along the Mississippi River. The Nutrient Reduction Strategy called for a 45% reduction in annual nitrogen. The America's Watershed Initiative's 2020 report card for the Mississippi River Basin (https://americaswatershed.org/reportcard/) rated the river's water quality as very poor, with conditions worsening with respect to their 2015 report card. Despite these major warnings, nitrate loads have increased in recent years. For instance, according to the Iowa Nutrient Reduction Strategy (https://store.extension.iastate.edu/product/15915) the nitrate loads between 2006 and 2010 have increased by ∼5% compared to the 1980–1986 period. Jones et al. (2018) analyzed the nitrate loads leaving Iowa between 1999 and 2016, and found that they have been increasing over the study period. Moreover, they did not attribute these changes to discharge or cropping intensity, and mentioned tile drainage as a potential factor. As we look into the future, Zhang et al. (2022) found that nitrate loads are projected to increase. Assuming that agricultural and management practices remain the same, they project an increase up to 30% in nitrogen loading, and attributed about half of this increase to heavy precipitation.

Issues related to water quality are just one side of the coin. The central United States is an area of the country that has been significantly affected by flooding as well, with the 2019 Missouri-Mississippi flood being the latest major event affecting the region. These events cause numerous fatalities and claim a large economic toll. Because of its impacts, many studies have examined the changes in the magnitude of annual maximum flood peaks. For instance, Villarini et al. (2011) examined trends in annual flood peaks at gauges with 75 years of record for the Midwest (16 in Iowa), finding that it was possible to detect abrupt or gradual changes only at a handful of sites. Archfield et al. (2016) examined the multidimensional behavior of flood changes for stream gauges across the United States, finding limited evidence for changes in the flood frequency, duration, peak magnitude, and volume in Iowa. Another important characteristic of annual maximum flood peaks is the frequency of events, for which there is stronger evidence of change. For example, Mallakpour and Villarini (2015) examined the frequency of annual maximum daily floods from 1962 to 2011 using a peaks-over-threshold (POT) approach. They found over 20 gauges in Iowa with a significant trend in the frequency of flood events. Slater and Villarini (2016) analyzed the number of days above the action and minor-stage flood level established by the National Weather Service and found increasing trends. Neri et al. (2019) showed that there are increasing trends in the number of flood events in the spring, summer, and fall seasons.

Intuitively, given that nitrate loads represent the product of nitrate concentration and discharge, it is reasonable to assume that nitrate loads are large when discharge is large, as is the case during floods. Hence, we can hypothesize that flooding and nitrate loads represent compound hazards, and they should be studied together rather than in isolation, evaluating the risks of co-occurring extremes of water quantity and quality. One key approach for analyzing the joint risk of compound events is through copulas. Copulas have been widely used in hydrology because of their flexibility (Chen & Guo, 2019; Genest & Favre, 2007; Salvadori & De Michele, 2007; Zhou et al., 2021), and represent the modeling tool used here. Therefore, the goal of this study is to model the dependence between discharge and nitrate loads when at least one is extreme in different agricultural watersheds in Iowa,

Core Ideas

- The bivariate models using nitrate loads and discharge across different sites in the state of Iowa are developed.
- Joint annual exceedance probabilities of two key events when at least one is extreme are estimated.
- The largest variability in highly likely values tends to be associated with high agricultural intensity.

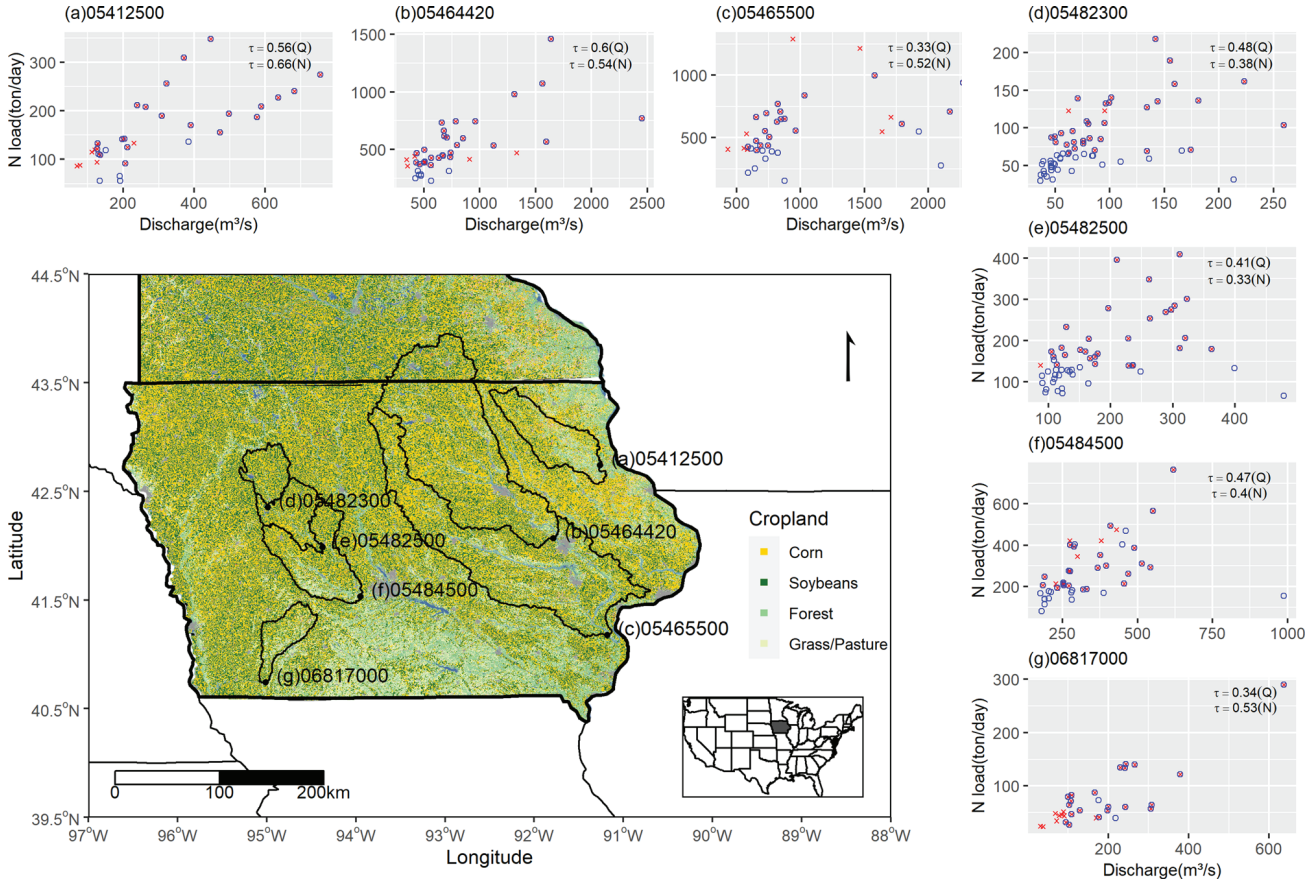

FIGURE 1 Location of the seven USGS stations and their basin boundaries, with information about the corn and soybean in 2020 as background (https://nassgeodata.gmu.edu/CropScape/). The scatterplots show the results based on the two samples conditioning on discharge (blue circles) and N loads (red crosses). The values in each scatterplot represent the Kendall's tau coefficient depending on whether N loads or discharge are the conditioning variable.

estimating their joint annual exceedance probabilities (AEPs) using bivariate copulas.

The rest of this paper is organized as follows. We describe data and methodology in the next section. Then, we present our results and discussion in Section 3. Section 4 summarizes the main findings and concludes the study.

2 | DATA AND METHODOLOGY

This study focuses on seven USGS sites in Iowa (Figure 1). We select these locations because they satisfied two conditions (1) the measurements of both daily mean discharge and daily mean nitrate plus nitrite ($NO_3 + NO_2$) loads (we will refer to it as "N loads") are available; and (2) the period of record for daily N loads is longer than 10 years. There are some gaps in the daily N loads time-series data (Figure 2). In this study, we calculate the daily N loads by multiplying N concentrations by daily discharge (units:ton/day). Furthermore, we select the years in which the completeness of the data records for both daily discharge and N load are over 70% during the March–November months (i.e., we excluded the winter season because of frozen conditions and extremely limited sampling during these months). Table 1 presents basic information of the sites and basins used in this study. The catchments are a few thousand square kilometers in size, with two of them almost an order of magnitude larger than the rest (i.e., 16,425 and 32,374 km^2). Agriculture is very intensive in this area, with the percentage of the basins cultivated in corn and soybean ranging between 55% and 82%. There is also a large number of hogs, generally part of animal feeding operations or concentrated animal feeding operations, and large expenses in fertilizers, all pointing to large amounts of nitrates applied and available in these basins.

In this study, we use a POT method that selects events above a sufficiently high threshold (Smith, 1984). The advantage of using a POT approach over an annual-maximum one is that it allows us to select the multiple extremes in the same block regardless of the year in which they happened (Santos et al., 2021). We apply a two-sided conditional sampling method that can address an asymmetric problem as extreme discharges and extreme N loads do not have to happen concurrently. Accordingly, we have two types of POT events to select, depending on whether N loads or discharge are the

TABLE 1 Basic information about the sites and basins considered in this study.

Site number	Site name	Latitude	Longitude	Period of record		Drainage area (km^2)	Type of commodity				
				Discharge	N load		Soybean (%)	Corn (%)	Soybean + corn (%)	Hogs (no./km^2)	Fertilizer (US dollars/km^2)
05412500	Turkey River at Garber	42.74	−91.26	08/08/1913–12/31/2021	05/09/2012–12/31/2021	4001	19.54	35.82	55.36	149	$13,258.81
05464420	Cedar River at Blairs Ferry Road at Palo	42.07	−91.79	03/25/2009–12/31/2021	10/14/2012–12/31/2021	16,425	30.14	42.86	73.00	149	$15,767.92
05465500	Iowa River at Wapello	41.18	−91.18	10/01/1914–12/31/2021	06/03/2009–12/31/2021	32,374	29.00	39.93	68.94	172	$14,606.32
05482300	North Raccoon River near Sac City	42.35	−94.99	06/01/1958–12/31/2021	03/27/2008–12/31/2021	1822	37.98	44.04	82.02	252	$14,874.68
05482500	North Raccoon River near Jefferson	41.99	−94.38	03/01/1940–12/31/2021	04/03/2008–12/31/2021	4193	37.24	45.16	82.40	260	$15,265.82
05484500	Raccoon River at Van Meter	41.53	−93.95	04/25/1915–12/31/2021	03/02/2012–12/31/2021	8912	32.48	40.75	73.23	216	$13,932.36
06817000	Nodaway River at Clarinda	40.74	−95.01	05/17/1918–12/31/2021	05/31/2012–12/31/2021	1973	30.45	30.97	61.42	32	$9839.14

Note: The results for the commodities are based on https://quickstats.nass.usda.gov/ and refer to the 2018 (2017) period for corn and soybeans (hogs and fertilizer).

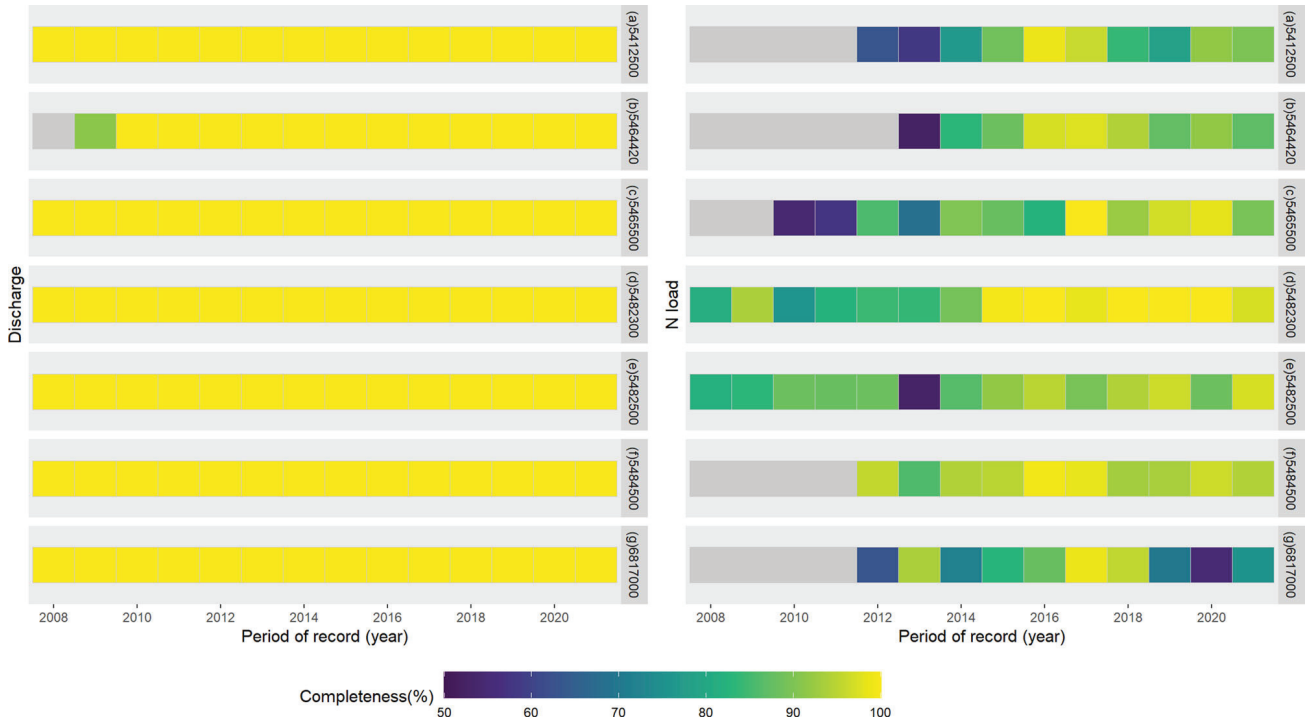

FIGURE 2 Completeness of data record for daily discharge (left panel) and N load (right panel) at the seven selected USGS sites (see their location in Figure 1).

conditioning variables (Bender et al., 2016; Jane et al., 2020; Kim et al., 2022). For the conditioning variable, we first select the POT events by setting a threshold that returns five events per year on average and allows only one peak in a ± 5-day window to ensure independence of the POT events (Lang et al., 1999). Then, we also select the concurrent events as the largest daily value of non-conditioning variable (i.e., N loads or discharge depending on whether discharge and N loads are the conditioning variables, respectively) within a 7-day window of selected POT events.

To model the bivariate relation between two variables (i.e., peak discharge and concurrent N loads, or peak N loads and concurrent discharge), we use copulas and the framework described by Jane et al. (2020). A copula C is a multivariate cumulative distribution function (CDF) with uniform distributed marginals in [0,1] (Joe, 1997). Sklar's theorem (Sklar, 1959) states that the joint CDF H of two variables X and Y, with marginal CDFs F and G, can be expressed as follows:

$$H(x, y) = C[F(x), G(y)] \quad (1)$$

where if F and G are continuous, then C is unique. Any bivariate joint distribution can thus be decomposed into a copula and the two marginal distributions, enabling the dependence between the variables to be modeled, through a copula, independently from their marginal distributions. Copulas consequently offer more flexibility than traditional bivariate modeling approaches where the joint distribution often constrains the choice of marginal distributions.

Parametric (or theoretical) copula families are copulas that can be expressed explicitly using one or more parameters. When combined with parametric marginal distributions, parametric copula families yield a fully parametric joint CDF, and thus provide a convenient way of describing the dependence between a pair of variables. Essentially the copula family controls the form of the dependence while the parameter(s) reflect the strength of the dependence. The rich array of bivariate copula families can represent a range of correlations, symmetries, and tail dependence, that is, the tendency for the extremes to coincide (Genest & Favre, 2007). The parameter(s) is typically estimated either directly from a dependence measure such as Kendall's τ (if a theoretical relationship exists) or through inference based on the true empirical copula (Sadegh et al., 2017). Early applications of copulas in hydrology primarily concerned flood frequency analysis (e.g., De Michele & Salvadori, 2003; Favre et al., 2004). Since then, bivariate copulas have gone on to account for spatial correlations of extremes at neighboring sites (e.g., Bender et al., 2016), been used in stochastic design storm generators (e.g., Kim et al., 2022; Vandenberghe et al., 2010) among other applications encompassing a variety of hydroclimatic variables (Tootoonchi et al., 2022). A flurry of recent studies chose copulas to model the joint probabilities of co-occurring high freshwater fluxes (rainfall/river discharge) and high sea levels (e.g., Wahl et al., 2015); this is motivated by impactful

events such as Hurricane Sandy (2012) and Hurricane Harvey (2017), where the interaction of these drivers likely exacerbated flooding. For more information on copulas, the reader is referred to Nelsen (2007).

Here we select the marginal distributions and apply the copula modeling based on the following steps regardless of the conditioning variable. The first step is to fit the marginal distributions to the POT values of the conditioning variable and the corresponding conditioned variable. For the POT events, we use the generalized Pareto distribution as it is the appropriate distribution when dealing with exceedances of a high threshold. For the conditioned variable, we select an appropriate distribution among 12 possible candidates (i.e., Birnbaum-Saunders, exponential, gamma, lognormal, normal, Tweedie, Weibull, logistic, Laplace, Gumbel, reversed Gumbel, and generalized gamma) with respect to the Akaike information criterion (AIC; Akaike, 1978). After estimating the marginal distributions, we consider 40 possible copula models (independence, Gaussian, Student *t*, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, BB8, Survival Clayton, Survival Gumbel, Survival Joe, Survival BB1, Survival BB6, Survival BB7, Survival BB8, rotated Clayton (90, 270˚), rotated Gumbel (90, 270˚), rotated Joe (90, 270˚), rotated BB1 (90, 270˚), rotated BB6 (90, 270˚), rotated BB7 (90, 270˚), rotated BB8 (90, 270˚), Tawn type 1, rotated Tawn type 1 (90, 180, 270˚), Tawn type 2, rotated Tawn type 2 (90, 180, 270˚)) and select the "best" one based on AIC. Consequently, we obtain two bivariate models for each station, which we bring together using the methodology described by Bender et al. (2016) to estimate the isolines for different joint AEPs. Given the two isolines (i.e., one from each copula model), we overlap them and take their outer envelope. To estimate the most probable event, we first generate 1,000,000 sample pairs of two samples (i.e., extreme discharge and N loads or extreme N loads and discharge) from the fitted models through Monte Carlo simulations; we then calculate the relative probability of events along the isoline using a kernel density estimate. Finally, we identify the most probable event on the combined isoline based on the relative probability of each point. A detailed description of combining two isolines is provided by Bender et al. (2016) and Jane et al. (2020). We assume stationary in our statistical modeling.

3 | RESULTS AND DISCUSSION

Figure 1 shows the statistical correlation between discharge and N load events for each site depending on which one is the conditioning variable, hence there are two Kendall's τ values. Based on Kendall's τ, the discharge and N loads are correlated for each of the basins, with correlation coefficients ranging from 0.33 to 0.66. All of Kendall's τ correlations are significant at the 5% level, pointing to the need to treat these events combined, rather than in isolation from each other. Moreover, many of the POT events are the same regardless of the conditioning variable, indicating that extreme discharge and extreme N loads tend to occur together.

To develop a bivariate model, we need to estimate the marginal distribution of each variable and a copula. Figures 3 and 4 show the results from fitting the marginal distributions for the samples conditioned on discharge and N loads for each of the seven sites. The generalized Pareto distribution can describe well the conditioning variable (left columns in Figures 3 and 4) across the different sites. For the non-conditioning variable, we consider 12 possible distributions and select the best one with respect to AIC. As summarized in Table 2, the gamma and lognormal distributions are selected most frequently as marginal distributions for the N loads, and the generalized gamma (gamma3) and lognormal distributions are selected most frequently as marginal distributions for discharge. Moreover, as shown in the middle column of Figures 3 and 4, the fitted distributions can accurately represent the observations. The results from fitting all 12 marginal distributions are shown in Figures S1 and S2.

In addition to the marginal distributions, we need to select a copula model, and we use AIC as selection criterion. As shown in Table 2, there is not a single model that is frequently selected, suggesting that the dependence structure between the two variables is not homogeneous across the state: this variability in dependence is visualized in the right column of Figures 3 and 4, where different sites exhibit different bivariate relationships requiring different copulas. Despite this variability, the selected copulas reproduce the joint dependence between discharge and N loads.

Once we have identified the different distributions and copulas, we can use this information to estimate the isolines for various AEPs, as shown in Figure 5 for return periods of 2, 5, 10, 25, 50, and 100 years (i.e., 0.5, 0.2, 0.1, 0.04, 0.02, and 0.01 AEPs) at the seven sites. These curves represent the outer envelope of the isolines derived from the two conditional samples. As expected from the different bivariate models, each location has a different structure. If we use the relative probability of occurrence of a given pair for the combined isoline, USGS 05412500 exhibits a tight range with a very high density (Figure 5a) compared to the other locations. One possible reason for this behavior has to do with the selected copulas, for which most of the mass is concentrated along a narrow region. Moreover, compared to the other basins, the percentage of the basin cultivated in corn and soybean for USGS 05412500 is the smallest among the selected watersheds, together with some of the smallest expenses in fertilizer and hog density.

The other basins have a much wider range of highly likely values (Figure 5). Compared to USGS 05412500, these basins are characterized by higher agricultural intensity, higher expenses in terms of fertilizers, and generally higher hog density. When we look at the three nested sites USGS 05482300,

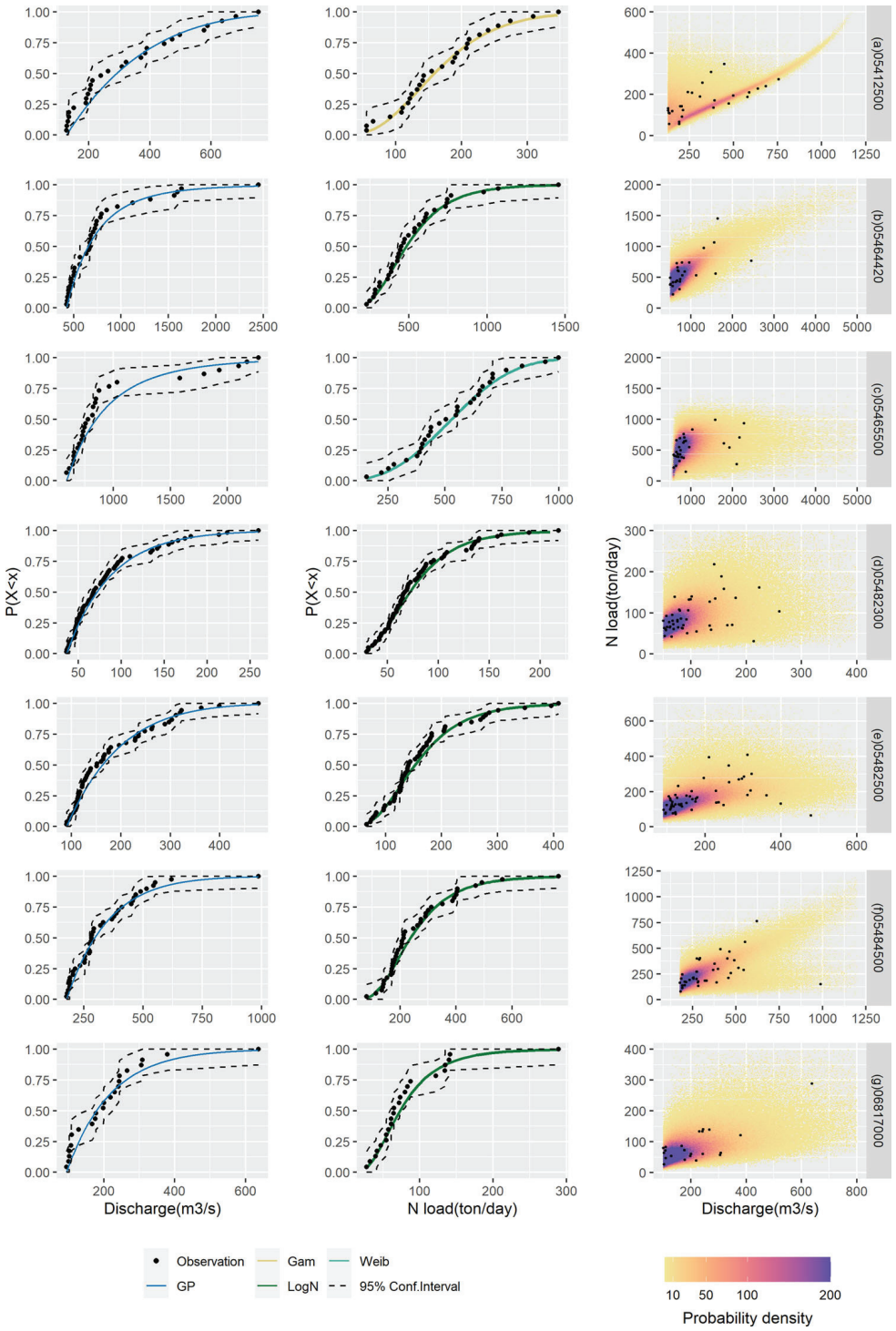

FIGURE 3 Left column: Results of fitting the POT discharge events with a generalized Pareto distribution (GP). Middle column: Results of fitting the N loads with several distributions (Gam, gamma distribution; LogN, lognormal distribution; Weib, Weibull distribution). Right column: Results of transformed realizations to the original scale based on the best-fitting copula. Here, the realizations are transformed to original scale by using inverse quantile functions of the selected marginal distributions. In the first two columns, the dashed lines represent the 95% confidence intervals. In all the panels, the black circles are the observations. Each row provides the results for each USGS station.

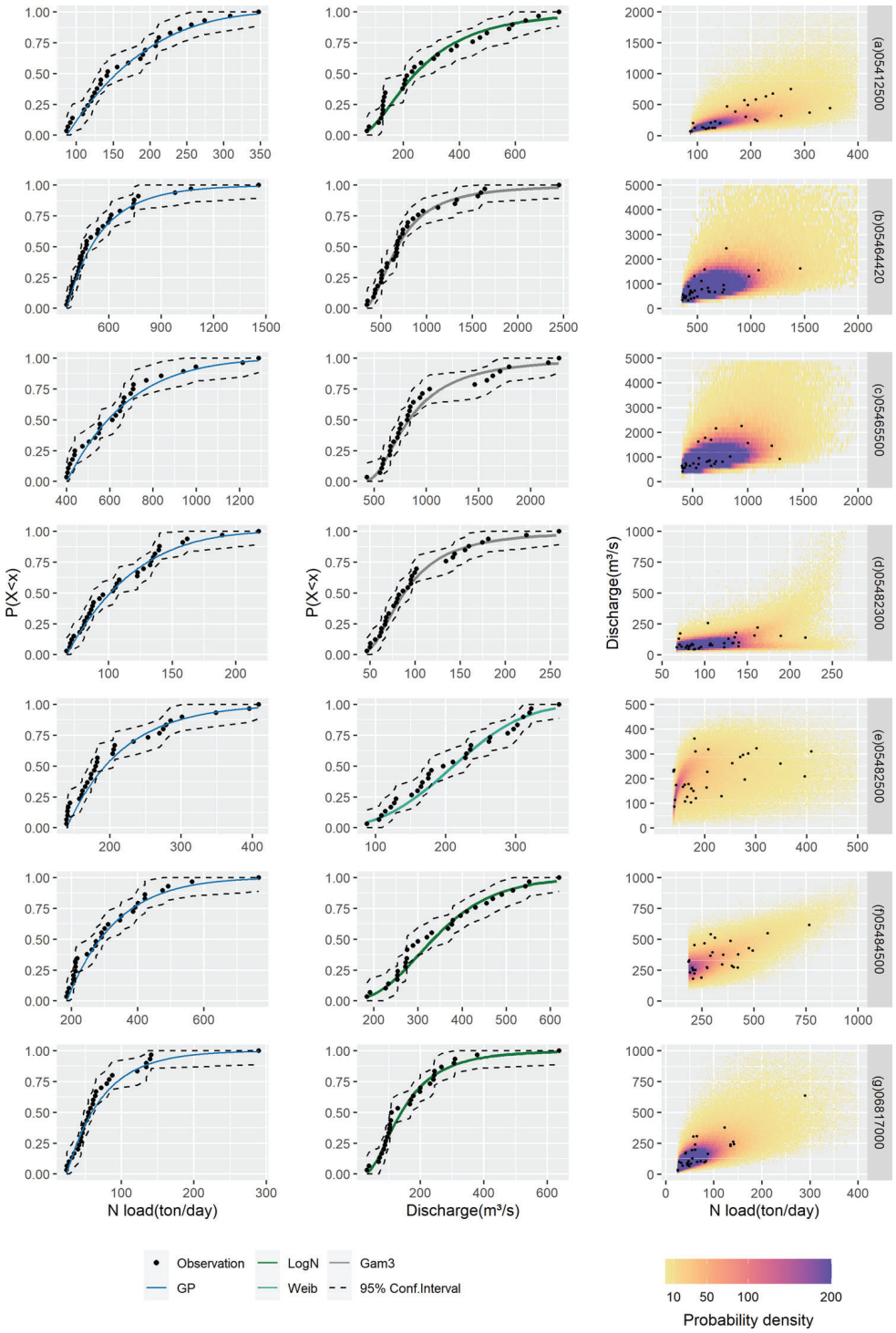

FIGURE 4 Left column: Results of fitting the POT N loads events with a generalized Pareto distribution (GP). Middle column: Results of fitting the discharge with several distributions (LogN, lognormal distribution; Weib, Weibull distribution; Gam3, generalized gamma distribution). Right column: Results of transformed realizations to the original scale based on the best-fitting copula. Here, the realizations are transformed to original scale by using inverse quantile functions of the selected marginal distributions. In the first two columns, the dashed lines represent the 95% confidence intervals. In all the panels, the black circles are the observations. Each row provides the results for each USGS station. POT, peaks-over-threshold.

TABLE 2 Summary of the selected marginal distributions and copula models for each site.

	Conditioning on discharge			Conditioning on N loads		
Site number	Conditioning data	Non-conditioning data	Copula model (AIC)	Conditioning data	Non-conditioning data	Copula model (AIC)
05412500	Generalized Pareto	Gamma	Tawn type 2 (−24.42)	Generalized Pareto	Lognormal	Survival Gumbel (−30.67)
05464420		Lognormal	Gumbel (−28.66)		Gamma3	Gaussian (−24.72)
05465500		Weibull	Rotated Tawn type 2 180° (−6.54)		Gamma3	Gaussian (−15.25)
05482300		Lognormal	Rotated Tawn type 2 180° (−40.61)		Gamma3	Tawn type 2 (−7.78)
05482500		Lognormal	Frank (−19.20)		Weibull	Rotated Tawn type 1180° (−6.48)
05484500		Lognormal	Tawn type 1 (−20.81)		Lognormal	Survival Clayton (−12.28)
06817000		Lognormal	Gaussian (−6.39)		Lognormal	Gaussian (−22.45)

Abbreviation: AIC, Akaike information criterion.

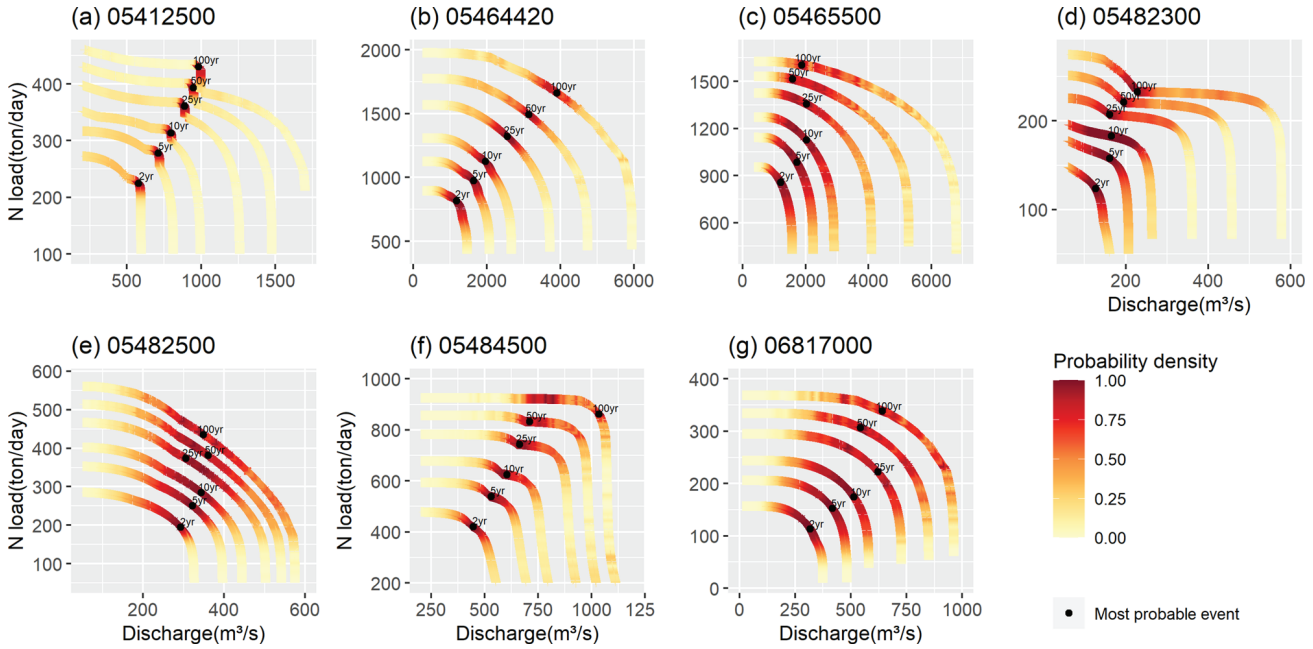

FIGURE 5 Quantile isolines and most probable events (black dots) for return periods of 2, 5, 10, 25, 50, and 100 years (i.e., 0.5, 0.2, 0.1, 0.04, 0.02, and 0.01 AEPs) at the seven sites in Figure 1.

05482500, and 05484500, the range with a high probability density is narrower for USGS 05484500, which is the largest basin and the most downstream station. It has the lowest percentage of harvested corn and soybeans acreage (i.e., 73% vs. 82%), the lowest hog density (i.e., 216 vs. 252 and 260 hogs/km^2) and the lowest amount of dollars spent in fertilizers (i.e., \$14 M/km^2 vs. ~\$15 M/km^2). Thus, one possible way to interpret these results is that, all else equal, there is a broader range of highly likely discharge and N loads values for basins that have been subject to intensive agriculture, compared to those who have been exporting lower N loads.

There is also variability in terms of the dependence of the most probable bivariate event on discharge values. For instance, for USGS 05412500, 05464420, 05484500, and 06817000, the most probable events tend to align on a sloped line for increasing AEPs, pointing to their dependence on both N loads and discharge. On the other hand, for USGS 05465500, 0548230, and 05482500, much of the variability in the most probable AEP values is driven by N loads, with discharge values around 2000 m^3/s, 200 m^3/s, and 300 m^3/s. While it is not immediately possible to come up with a definitive reason for these differences, one element that these three sites have in common is that they are basins that have been subject to high agricultural intensity, a high density of hogs, and large expenditures in terms of fertilizer.

The results in Figure 5 are based on using the "best" marginal distribution. Here, we also explore the sensitivity of the results to the selection of marginal distribution, comparing the results in Figure 5 against those obtained from the second- and third-best marginal distributions. The second- and third-best marginal distributions for discharge and N loads fit well the observations, with small differences in performance compared to the best one according to the AIC (see Figures S3 and S4). The combined isolines show that there are small differences up to 0.1 AEPs, which become larger for smaller AEPs (Figure S5). Moreover, the combined isolines with the best marginal distributions mostly estimates larger discharge and N loads for a given AEP, resulting in more conservative design values in light of the observed compounding. One element worth highlighting is that the largest differences among isolines tend to occur when the generalized gamma distribution (Gam3) is selected for discharge; this is a distribution with a heavier tail than the others, leading to generally larger events for low AEPs.

4 | CONCLUSIONS

In this study, we developed bivariate models to describe the compound effects of extreme discharge and extreme N loads at seven locations across the state of Iowa. The main conclusions of our work can be summarized as follows:

- The generalized Pareto distribution describes well the observational peak discharge and N loads records, while the lognormal, Weibull, generalized gamma, and gamma distributions are appropriate for the coinciding N loads and discharge. In terms of copula models, the selected copula varies across sites reflecting the diverse nature of bivariate dependence across the sites.

- The joint AEPs exhibit different dependence structures, highlighting differences in the controlling roles of discharge and N loads.
- We examined the role of agricultural intensity (i.e., harvested corn and soybean acreage as a percentage of the entire watershed), hog density (i.e., proxy for manure, hence nitrate), and fertilizer expenditures in explaining the variability in results. We found that the basins for which there was a broad range of highly likely values for a given AEP were also some of the most intensively farmed ones, with high hog density and fertilizer expenditures.
- Different marginal distributions tend to describe well the observations. We show that the isolines tend to be sensitive to the choice of marginal distribution, especially for low AEPs due to the large sampling uncertainties and the relatively short records considered here.

The results of our study about water quantity and quality expand our understanding of the nature of compound extremes to a topic that has received limited attention in the literature. While our emphasis was on N loads at seven sites in Iowa, the methodology and approach are general and could be extended to other sites for which long-term records are available as well as to other pollutants.

We provide an initial assessment of the potential drivers that could explain the variability in results among the different sites. However, more work needs to be performed to be able to attribute more definitively the observed variability to different drivers (e.g., average width of the riparian buffer). This could be aided by considering a broader set of stations even beyond Iowa. With that said, one potential issue with a future study of this kind is that it would require locations with long term-records. There is limited availability of continuous nitrate data not just in Iowa but across the contiguous United States. Our capability of understanding the connections between water quantity and quality in a holistic way resides in availability of high-quality and long-term records, and the prioritization of limited resources towards this goal.

AUTHOR CONTRIBUTIONS

Taereem Kim: Data curation; investigation; validation; writing–original draft; writing–review and editing. **Gabriele Villarini**: Conceptualization; funding acquisition; investigation; supervision; writing–review and editing. **Hanbeen Kim**: Data curation; formal analysis; methodology; writing–review and editing. **Robert Jane**: Formal analysis; methodology; software; writing–review and editing. **Thomas Wahl**: Formal analysis; methodology; software; writing–review and editing.

ACKNOWLEDGMENTS

The authors acknowledge support from IIHR—Hydroscience & Engineering, and the USACE Climate Preparedness and Resilience Community of Practice and Programs. T. W. was supported by the National Science Foundation under Grant AGS-192938. The authors are thankful to the comments by Dr. Hirsch and an anonymous reviewer.

Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Taereem Kim https://orcid.org/0000-0001-8689-1112

REFERENCES

Akaike, H. (1978). On the likelihood of a time series model. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *27*(3–4), 217–235.

Archfield, S. A., Hirsch, R. M., Viglione, A., & Blöschl, G. (2016). Fragmented patterns of flood change across the United States. *Geophysical Research Letters*, *43*(19), 10,232–210,239. https://doi.org/10.1002/2016GL070590

Bender, J., Wahl, T., Müller, A., & Jensen, J. (2016). A multivariate design framework for river confluences. *Hydrological Sciences Journal*, *61*(3), 471–482. https://doi.org/10.1080/02626667.2015.1052816

Chen, L., & Guo, S. (2019). *Copulas and its application in hydrology and water resources*. Springer.

De Michele, C., & Salvadori, G. (2003). A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research: Atmospheres*, *108*(D2), 4067–4078. https://doi.org/10.1029/2002JD002534

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, *40*(1). https://doi.org/10.1029/2003WR002456

Genest, C., & Favre, A. -C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, *12*(4), 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Jane, R., Cadavid, L., Obeysekera, J., & Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in south Florida. *Natural Hazards and Earth System Sciences*, *20*(10), 2681–2699. https://doi.org/10.5194/nhess-20-2681-2020

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC press.

Jones, C. S., Nielsen, J. K., Schilling, K. E., & Weber, L. J. (2018). Iowa stream nitrate and the Gulf of Mexico. *PLoS ONE*, *13*(4), e0195930. https://doi.org/10.1371/journal.pone.0195930

Kim, H., Villarini, G., Jane, R., Wahl, T., Misra, S., & Michalek, A. (2022). On the generation of high-resolution probabilistic design events capturing the joint occurrence of rainfall and storm surge in coastal basins. *International Journal of Climatology*, *42*(2), 761–771. https://doi.org/10.1002/joc.7825

Lang, M., Ouarda, T. B. M. J., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, *225*(3–4), 103–117. https://doi.org/10.1016/S0022-1694(99)00167-5

Liu, Y., Wang, J., Cao, S., Han, B., Liu, S., & Chen, D. (2022). Copula-based framework for integrated evaluation of water quality and quantity: A case study of Yihe river, China. *Science of The Total Environment*, *804*, 150075. https://doi.org/10.1016/j.scitotenv.2021.150075

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*(3), 250–254. https://doi.org/10.1038/nclimate2516

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D., & Vrac, M. (2018). Soil moisture drought in Europe: A compound event of precipitation and potential evapotranspiration on multiple time scales. *Journal of Hydrometeorology*, *19*(8), 1255–1271. https://doi.org/10.1175/JHM-D-18-0017.1

Mueller, D. K., & Helsel, D. R. (1996). *Nutrients in the nation's waters: Too much of a good thing?* (Vol. 1136). US Government Printing Office.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Neri, A., Villarini, G., Slater, L. J., & Napolitano, F. (2019). On the statistical attribution of the frequency of flood events across the US Midwest. *Advances in Water Resources*, *127*, 225–236. https://doi.org/10.1016/j.advwatres.2019.03.019

Oeurng, C., Sauvage, S., & Sánchez-Pérez, J. -M. (2010). Temporal variability of nitrate transport through hydrological response during flood events within a large agricultural catchment in south-west France. *Science of The Total Environment*, *409*(1), 140–149. https://doi.org/10.1016/j.scitotenv.2010.09.006

Ribeiro, A. F. S., Russo, A., Gouveia, C. M., & Páscoa, P. (2019). Copula-based agricultural drought risk of rainfed cropping systems. *Agricultural Water Management*, *223*, 105689. https://doi.org/10.1016/j.agwat.2019.105689

Sadegh, M., Ragno, E., & AghaKouchak, A. (2017). Multivariate copula analysis toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework. *Water Resources Research*, *53*(6), 5166–5183. https://doi.org/10.1002/2016WR020242

Salvadori, G., & De Michele, C. (2007). On the use of copulas in hydrology: Theory and practice. *Journal of Hydrologic Engineering*, *12*(4), 369–380. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(369)

Santos, V. M., Wahl, T., Jane, R., Misra, S. K., & White, K. D. (2021). Assessing compound flooding potential with multivariate statistical models in a complex estuarine system under data constraints. *Journal of Flood Risk Management*, *14*(4), e12749. https://doi.org/10.1111/jfr3.12749

Sebastian, A., Dupuits, E. J. C., & Morales-Nápoles, O. (2017). Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed. *Coastal Engineering*, *125*, 42–50. https://doi.org/10.1016/j.coastaleng.2017.03.008

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, *8*, 229–231.

Slater, L. J., & Villarini, G. (2016). Recent trends in US flood risk. *Geophysical Research Letters*, *43*(24), 12,428–412,436. https://doi.org/10.1002/2016GL071199

Smith, R. L. (1984). Threshold methods for sample extremes. In *Statistical extremes and applications* (pp. 621–638). Springer.

Tilloy, A., Malamud, B., & Joly-Laugel, A. (2021). A methodology for the spatiotemporal identification of compound hazards: Wind and precipitation extremes in Great Britain (1979–2019). *Earth System Dynamics Discussions*. https://doi.org/10.5194/esd-2021-52

Tilloy, A., Malamud, B. D., Winter, H., & Joly-Laugel, A. (2020). Evaluating the efficacy of bivariate extreme modelling approaches for multi-hazard scenarios. *Natural Hazards and Earth System Sciences*, *20*(8), 2091–2117. https://doi.org/10.5194/nhess-20-2091-2020

Tootoonchi, F., Sadegh, M., Haerter, J. O., Räty, O., Grabs, T., & Teutschbein, C. (2022). Copulas for hydroclimatic analysis: A practice-oriented overview. *Wiley Interdisciplinary Reviews: Water*, *9*(2), e1579.

Turner, R. E., & Rabalais, N. N. (1994). Coastal eutrophication near the Mississippi river delta. *Nature*, *368*(6472), 619–621. https://doi.org/10.1038/368619a0

Vandenberghe, S., Verhoest, N., Buyse, E., & De Baets, B. (2010). A stochastic design rainfall generator based on copulas and mass curves. *Hydrology and Earth System Sciences*, *14*(12), 2429–2442. https://doi.org/10.5194/hess-14-2429-2010

Villarini, G., Smith, J. A., Baeck, M. L., & Krajewski, W. F. (2011). Examining flood frequency distributions in the Midwest U.S. *Journal of the American Water Resources Association*, *47*(3), 447–463. https://doi.org/10.1111/j.1752-1688.2011.00540.x

Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, *5*(12), 1093–1097. https://doi.org/10.1038/nclimate2736

Wang, X., Zang, N., Liang, P., Cai, Y., Li, C., & Yang, Z. (2017). Identifying priority management intervals of discharge and TN/TP concentration with copula analysis for Miyun Reservoir inflows, North China. *Science of The Total Environment*, *609*, 1258–1269. https://doi.org/10.1016/j.scitotenv.2017.07.135

Ward, M. H., Jones, R. R., Brender, J. D., De Kok, T. M., Weyer, P. J., Nolan, B. T., Villanueva, C. M., & Van Breda, S. G. (2018). Drinking water nitrate and human health: An updated review. *International Journal of Environmental Research and Public Health*, *15*(7), 1557. https://doi.org/10.3390/ijerph15071557

Zellou, B., & Rahali, H. (2019). Assessment of the joint impact of extreme rainfall and storm surge on the risk of flooding in a coastal area. *Journal of Hydrology*, *569*, 647–665. https://doi.org/10.1016/j.jhydrol.2018.12.028

Zhang, J., Lu, C., Crumpton, W., Jones, C., Tian, H., Villarini, G., Schiling, K., & Green, D. (2022). Heavy precipitation impacts on nitrogen loading to the Gulf of Mexico in the 21st century: Model projections under future climate scenarios. *Earth's future*, *10*(4), e2021EF002141.

Zhou, Y., Zhang, D., Zhou, P., Wang, Z., Yang, P., Jin, J., Cui, Y., & Ning, S. (2021). Copula-Based bivariate return period analysis and its implication to hydrological design event. *Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.12943

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., & Mahecha, M. D. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, *1*(7), 333–347.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Kim, T., Villarini, G., Kim, H., Jane, R., & Wahl, T. (2023). On the compounding of nitrate loads and discharge. *Journal of Environmental Quality*, *52,* 706–717. https://doi.org/10.1002/jeq2.20458